Differing Air Traffic Controller Responses to Similar Trajectory Prediction Errors

An Interrupted Time-Series Analysis of Controller Behavior

Joey Mercer
Airspace Operations Laboratory
NASA Ames Research Center
Mountain View, CA, USA
joey.mercer@nasa.gov

Sarah Hunt Espinosa
User-Centered Design & Engineering
SPAWAR Systems Center Pacific
San Diego, CA, USA
shunt@spawar.navy.mil

Nancy Bienert, Sean Laraway, San José State University / NASA Ames Research Center, Moffett Field, CA

Abstract— **A Human-In-The-Loop simulation was conducted in January of 2013 in the Airspace Operations Laboratory at NASA's Ames Research Center. The simulation airspace included two en route sectors feeding the northwest corner of Atlanta's Terminal Radar Approach Control. The focus of this paper is on how uncertainties in the study's trajectory predictions impacted the controllers' ability to perform their duties. Of particular interest is how the controllers interacted with the delay information displayed in the meter list and data block while managing the arrival flows. Due to wind forecasts with 30-knot over-predictions and 30-knot under-predictions, delay value computations included errors of similar magnitude, albeit in opposite directions. However, when performing their duties in the presence of these errors, did the controllers issue clearances of similar magnitude, albeit in opposite directions?**

Keywords—Air Traffic Control, Human-in-the-Loop Simulation, Trajectory Prediction Uncertainty, Human-Automation Interaction, Interrupted Time-Series

I. INTRODUCTION

The National Airspace System (NAS) forecasts continued growth in traffic demand [1], and under the plans for the Next Generation Air Transportation System (NextGen), the Federal Aviation Administration (FAA) aims to address one of the system's constraining factors, the controller's mental capacity, by increasing the use of automation aids [2, 3]. These tools depend on the predicted speed and path of an aircraft: its trajectory. Trajectory prediction capabilities are therefore a fundamental part of future Air Traffic Management (ATM) systems, and will be used by NextGen automation tools to provide advisory aids. The performance of such Decision Support Tools (DSTs), and ultimately their operational acceptance, is likely dependent on the accuracy of the underlying trajectory predictions. However, trajectory predictions, by their very nature, are not perfect: they are informed guesses.

With the predicted trajectory of an aircraft, NextGen's advanced DSTs can provide an alert: either cautionary (such as highlighting a conflict) or informational (such as suggesting a speed to meet a schedule, or displaying the predicted delay at a metering point), if needed. Uncertainties in trajectory predictions then, directly affect the DST's ability to help the controller perform their task. While it is possible that the controller could compensate for the errors in 'bad automation,' they can only do so to a limited extent. More specifically, the controller's efforts to compensate for the automation may reach a workload ceiling, at which point performance may worsen.

A. Study of Trajectory Prediction Uncertainties

A Human-In-The-Loop simulation was conducted in January of 2013 [4] in the Airspace Operations Laboratory (AOL) at NASA's Ames Research Center [5]. The simulation airspace included two en route sectors (one high-altitude and one low-altitude) feeding the northwest meter-fix of Atlanta's Terminal Radar Approach Control (TRACON), depicted in Figure 1. The test participants were responsible for delivering aircraft to the meter-fix within +/- 20 seconds of the scheduled time (information shown to them on their display in a meter list and in aircraft data blocks), and for providing standard separation services for all aircraft. The environment also included over-flight and departure traffic, thereby increasing the complexity of the task of providing separation.

The participants staffing the test sectors were retired air traffic controllers, none of whom were familiar with the test airspace, and had an average of 23.75 years of experience and had been retired for an average of 5.5 years. Confederate controllers, also retired, staffed Radar-Associate (D-side) positions, one for each test sector, while student/general aviation pilots staffed the confederate pseudo-pilot positions. During a one-week study, two separate simulations were conducted simultaneously and in parallel, creating two 'worlds'. Although the approach of two worlds required twice

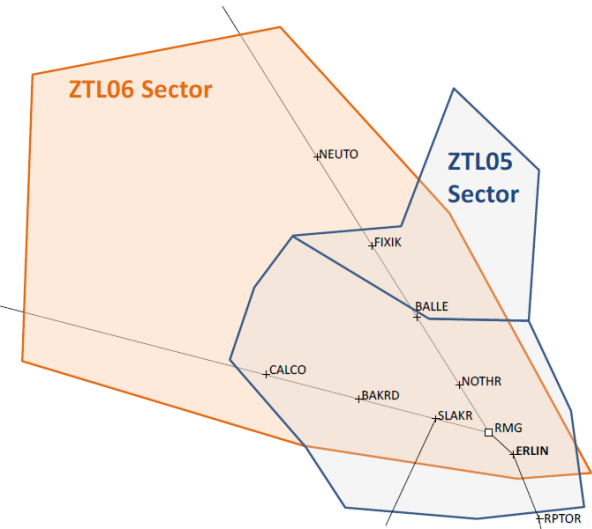the number of positions described above, it doubled the amount of data collected in the same amount of time.

Figure 1: Experimental sectors simulating an arrival flow into Atlanta.

The time-frame of the operations simulated in this study was anchored in a notional 'mid-term' NextGen environment, where better surveillance data and advanced DSTs could be expected. In performing their duties, the controllers had several automation aids available to them (see Figure 2). Meter lists contained arrival sequence and schedule information, while delay indications were also shown in the data block. A conflict probe displayed information regarding detected conflicts in a conflict list as well as in the data block. Trial-planning tools were also available, allowing controllers to manually craft trajectory changes to solve local problems. Additionally, functions were available to the controllers that allowed them to request a solution from the automation for detected problems.
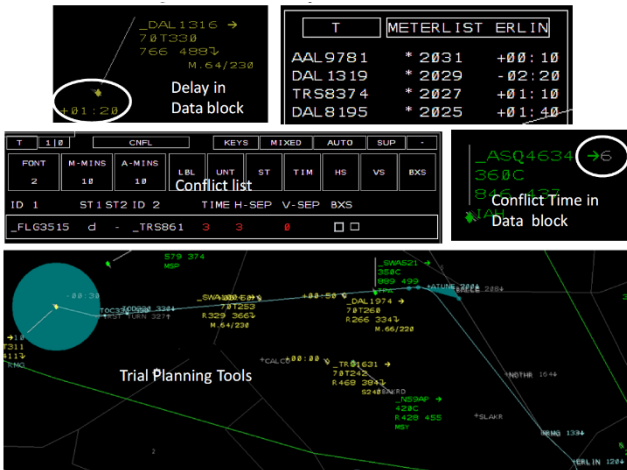

Figure 2: Screenshots of mid-term operational tools.

It is important to note that the delay information displayed in the meter list and data blocks was configured with a unique behavior that could be considered a tool in its own right. Any amendments made by the controllers to an aircraft's trajectory, either manually entered, or trial-plan-assisted, caused the ground system's automation to immediately compute a new trajectory which incorporated the newly available information. For example, if the controller issued a speed clearance to an aircraft, when inputting the new speed as a system entry, the automation would then compute a new trajectory for the aircraft based on the new speed. This had an immediate effect on the delay information displayed in the data block and meter list, which would immediately update to reflect the new trajectory prediction.

Because all tools were trajectory-based, they were subject to the various simulated errors inherent in those trajectories, meaning that due to uncertainties in the trajectories provided, the tool information displayed to the controllers was imperfect. This highlights one of the simulation's primary objectives: to examine at which point the automation tools would become unacceptable to the controllers and no longer support adequate system performance in terms of separation services or metering conformance.

The simulation investigated how trajectory prediction uncertainties impacted the controller's ability to provide standard separation services and deliver aircraft on time to the meter-fix. This paper aims to understand the nature of the controllers' response to different uncertainties simulated across similar conditions. Of interest is how the participants' form of compensating for prediction errors affected the clearances they issued, and the exploration of new analysis methods to gain further insight into controller behavior.

B. Simulation of Trajectory Prediction Errors

Uncertainties were introduced in the form of wind forecast errors and errors in aircraft performance assumptions (e.g., climb/descent rates). A selection of different Rapid-Update Cycle (RUC) wind files created mismatches between environment and forecast wind fields. Wind forecast errors either over-predicted or under-predicted a predominant tail wind by varying amounts. A baseline condition with no wind errors was included, as well as 'Realistic' Root-Mean-Square (RMS) wind errors of 10 knots, meant to represent typical 'real-world' forecast errors. Other levels of wind error included 'Moderate' RMS errors of 20 knots, and 'Large' RMS errors of 30 knots.

The simulation also investigated errors in the underlying aircraft performance models, which were implemented such that while the ground system's assumptions about aircraft performance remained constant, the actual descent and climb performance of individual aircraft behaved according to modified 'scale factors.' The scale factors were designed to impact the distance normally needed by an aircraft to descend from one altitude constraint to the next, or to climb from one altitude constraint to the next. The simulation examined a baseline condition with no aircraft performance errors, as well as two additional target levels of aircraft performance model

errors: 'Realistic' errors of approximately 5%, and 'Large' performance errors of approximately 25%.

The simulation employed two primary scenarios (scenarios A and B), designed independently, but meant to be comparable. Coupling the two scenarios with different combinations of forecast and environment (i.e., 'truth') winds allowed the simulation to not only examine different magnitudes of wind error, but also both directions of error bias. A positive bias (an over-prediction error) resulted when the forecast winds were stronger than the environment winds, whereas forecast winds that were weaker than environment winds resulted in a negative bias (an under-prediction error). Environment and forecast winds were paired in these two ways for each of the wind-forecast error conditions.

During the simulation, traffic scenario A was mostly paired with positive-bias wind errors, whereas traffic scenario B was mostly paired with negative-bias wind errors. The negative-bias wind errors used during trials with scenario B had the effect of presenting the controllers with seemingly smaller initial delay values. In contrast, the positive-bias wind errors used during trials with scenario A impacted the trajectory predictions such that the controllers saw seemingly larger initial delay values. For an aircraft left untouched, the delay would gradually correct towards the actual delay (i.e., the delay expected using perfect wind forecast information) as it came closer to the meter-fix. This is true in either scenario: that is, the seemingly smaller initial delay value in scenario B would gradually increase as the aircraft approached the meter-fix, and conversely, the seemingly larger initial delay value in scenario A would gradually decrease.

II. COMPARISONS MADE

The simulation results showed that scenario B was less challenging for the controllers than scenario A [4], [6]. Given the different wind-error biases associated with each scenario, the present investigation explores the relationship between the wind-error biases and the study's findings. To achieve this, runs 11 and 12 from the study are examined because they both simulated 30-knot forecast wind errors: run 11 did so with a positive wind-error bias, while run 12 did so with a negative wind-error bias.

Central to the current analysis is the controller's response to the trajectory prediction errors as a result of the wind-error bias. While it is true that the entire set of instructions issued by the controller to an aircraft represents that response, this analysis distinguishes the first speed clearance from the remaining clearances, thereby isolating what the authors believe best represents the controller's initial judgement of the response needed, from later corrective actions. In this regard, although the assigned speed is informative, it comes from a limited range of flyable speeds. Therefore, the current analyses consider the relative magnitude of the issued speed change (i.e., the difference between the aircraft's current speed and the issued speed), and when examined for all aircraft over the course of a run, provide insight in how the controller's judgement changed over time. This approach embraces the natural learning controllers do as they issue clearances, observe their effect, and adjust accordingly for the next clearance.

Exploring assigned speeds and speed-change magnitudes together allows for a multi-dimensional analysis of the controller's behavior not possible with just one metric.

The scenarios used in runs 11 and 12 were similar but not identical, yielding within-subject data that is not directly comparable and does not fit traditional statistical testing. Instead, an Interrupted Time-Series (ITS) analysis was used to detect any progressive changes over time for a given participant.

III. RESULTS

Primarily based on the work done by Huitema [7], the ITS used here is based on an A-B design and traditional linear regression. Like linear regression, each data set has an appropriate model. The available models for this analysis are listed in Figure 3 and will be referenced numerically (1-4).

<div style="border:1px solid">

Interrupted Time Series Models (Huitema, 2011)

Model 1: $Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 SC_t + \varepsilon_t$

Model 2: $Y_t = \beta_0 + \beta_1 D_t + \varepsilon_t$

Model 3: $Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 SC_t + \emptyset_1 \varepsilon_{t-1} + \mu_t$

Model 4: $Y_t = \beta_0 + \beta_1 D_t + \emptyset_1 \varepsilon_{t-1} + \mu_t$

</div>

Figure 3: Models for an A-B design Interrupted Time-Series Analysis (Huitema, 2011)

Model selection is based on, first, whether slope is a necessary variable to explain the data, and secondly, whether or not the errors are auto-regressive. If slope is a necessary variable, models 1 and 3 are candidates, if not, than models 2 and 4. If the errors are auto-regressive, then models 3 and 4 are the only choices because they contain the appropriate correction. These constraints lead to the selection of the appropriate model, and Huitema [7] lays out the necessary steps for model selection. This analysis uses models 1, 2, and 3. Besides slope, the primary variable of interest is level change. This is the magnitude of change, or level, between the predicted first point in run 12, assuming no change from run 11 (the baseline condition), and the actual first data point in run 12. This indicates the immediate effect of the treatment, in this case, the reversal of wind error, on behavior. If no slope is required, level change alone explains the data. However, the addition of slope (when possible) leads to further interpretation of the data.

A. World 1, sector 5

Table 1: World 1, Sector 5's results.

World 1, Sector 5	Speed Events	Speed Change Magnitude
Model	1	2
Level Change (D)	no	yes
Baseline Slope [Run 11] (T)	yes	--
Slope Change (SC)	yes	--
Autocorrelated?	no	no

W1S5 did not show a significant level change from runs 11 to 12 in their raw speed event data β = -3.58, $t(112)$ = .40, p =.69, SLC= .15, however the level change for speed-change magnitude was significantly different β = -1.04, $t(114)$ = -5.57, p <.001, SLC= -1.04. As seen in Figure 4, there was a significant slope in the baseline of speed events β = .47, $t(112)$ = 2.63, p= .001, SLC= .15 and significant speed event slope level change β = -.58, $t(112)$ = -2.15, p = 0.03, SLC= 0.15 from runs 11 to 12.
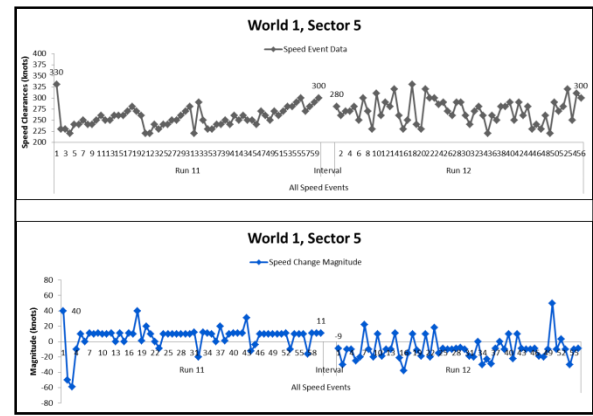

Figure 4: World 1, Sector 5's graphical display of the chronological data.

B. World 2, sector 5

Table 3: World 2, Sector 5's results.

Table 2: Statistical summary of the interrupted time series analysis.

Measure	Significance				Effect Size- R^2					Standardized Level Change Statistic		
	Beta	t	df	p	R^2	F	df1	df2	p	B_{lc}	MSresidual	SLC Statistic
World 1, Sector 5: Model 1: Speed Events												
Level Change	3.58	0.40	112.00	0.69	na	na	na	na	na	3.5819	581.2249	0.1486
Baseline Slope	0.47	2.63	112.00	0.01*								
Slope Change	-0.58	-2.15	112.00	0.034*								
World 1, Sector 5: Model 2: Speed Change Magnitude												
Level Change	-1.0351	-5.571	114	0.00**	0.2400	31.036	1	114	0.00	-15.6670	229.0731	-1.0351
Baseline Slope	na	na	na	na								
Slope Change	na	na	na	na								
World 2, Sector 5: Model 2: Speed Events												
Level Change	12.3915	2.083	60	0.041*	0.0670	4.341	1	60	0.041	12.3915	539.1922	0.5336
Baseline Slope	na	na	na	na								
Slope Change	na	na	na	na								
World 2, Sector 5: Model 1: Speed Change Magnitude												
Level Change	-40.8620	-4.552	58	0.00**	na	na	na	na	na	-40.8620	310.5021	-2.3189
Baseline Slope	0.17	0.577	58	0.566								
Slope Change	1.455	2.766	58	0.008*								
World 1, Sector 6: Model 2: Speed Events												
Level Change	47.8920	9.372	55	0.00**	0.6150	87.826	1	55.00	0.00	47.8920	339.0470	2.6010
Baseline Slope	na	na	na	na								
Slope Change	na	na	na	na								
World 1, Sector 6: Model 2: Speed Change Magnitude												
Level Change	36.9135	5.274	55	0.00**	0.3360	27.811	1	55.00	0.00	36.9135	636.0740	1.4636
Baseline Slope	na	na	na	na								
Slope Change	na	na	na	na								
World 2, Sector 6: Model 3: Speed Events												
Level Change	34.0865	2.023	59	0.0476*	na	na	na	na	na	34.0865	331.6938	1.8716
Baseline Slope	0.8017	1.127	59	0.2643								
Slope Change	-1.3509	-1.064	59	0.2917								
World 2, Sector 6: Model 3: Speed Change Magnitude												
Level Change	1.2879	0.098	59	0.9223	na	na	na	na	na	1.2879	426.9135	0.0623
Baseline Slope	1.6001	3.596	59	0.0007*								
Slope Change	-1.6282	-2.036	59	0.0462*								

World 2, Sector 5	Speed Events	Speed Change Magnitude
Model	2	1
Level Change (D)	yes	yes
Baseline Slope [Run 11] (T)	--	no
Slope Change (SC)	--	yes
Autocorrelated?	*no*	*no*

W2S5 had a statistically significant level change for both speed event data β = 12.39, t(60) = 2.083, p =0.04, SLC= 0.53 and speed-change magnitude β = -40.86, t(58) = -4.55, p <.001, SLC= -2.32, which can be seen in Figure 5. Model 2 fully explained the speed event data, removing slope. However, the speed-change magnitude data required model 1, and while there was no statistically significant slope in run 11, there was a statistically significant slope change from runs 11 to 12 β = 1.46, t(58) = 2.77, p = 0.008, SLC= -2.32
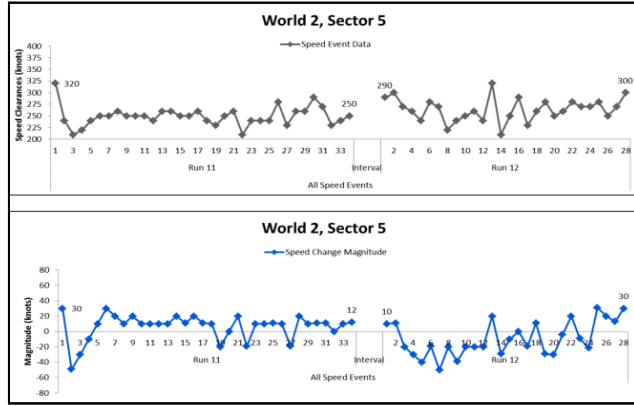

Figure 5: World 2, Sector 5's graphical display of the chronological data.

C. *World 1, sector 6*

Table 4: World 1, Sector 6's results.

World 1, Sector 6	Speed Events	Speed Change Magnitude
Model	2	2
Level Change (D)	yes	yes
Baseline Slope [Run 11] (T)	--	--
Slope Change (SC)	--	--
Autocorrelated?	*no*	*no*

W1S6 used model 2 to explain both data sets, with level change the only variable of interest (see Figure 6). For both speed events β = 47.89, t(55) = 9.37, p <.001, SLC= 2.60 and speed-change magnitude β = 36.91, t(55) = 5.27, p <.001, SLC= 1.44 there was a statistically significant level change between runs 11 and 12.
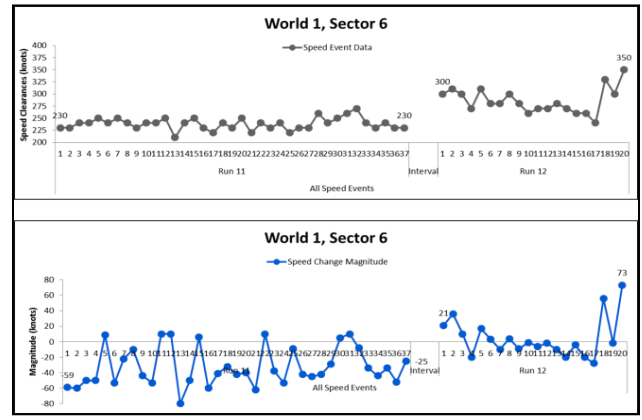

Figure 6: World 1, Sector 6's graphical display of the chronological data.

D. *World 2, sector 6*

Table 5: World 2, Sector 6's results.

World 2, Sector 6	Speed Events	Speed Change Magnitude
Model	3	3
Level Change (D)	yes	no
Baseline Slope [Run 11] (T)	no	yes
Slope Change (SC)	no	yes
Autocorrelated?	*yes*	*yes*

W2S6 had the only instance of auto-regressive errors, and with the correction, both data sets used model 3. Level change was statistically significant in speed event β = 34.09, t(59) = 2.02, p =0.05, SLC= 1.87 but not speed-change magnitude β = 1.29, t(59) = 0.01, p =0.92, SLC= 0.06. Speed event data did not have a statistically significant baseline slope β = 0.8, t(59) = 1.13, p =0.26, SLC= 1.87 or slope change β = -1.35, t(59) = -1.06, p =0.29, SLC= 1.87. However, speed-change magnitude had both a statistically significant baseline slope β = 1.60, t(59) = 3.6, p <.001, SLC= 0.06 and slope change β = -1.63, t(59) = -2.04, p =0.05, SLC= 0.06 (see Figure 7).
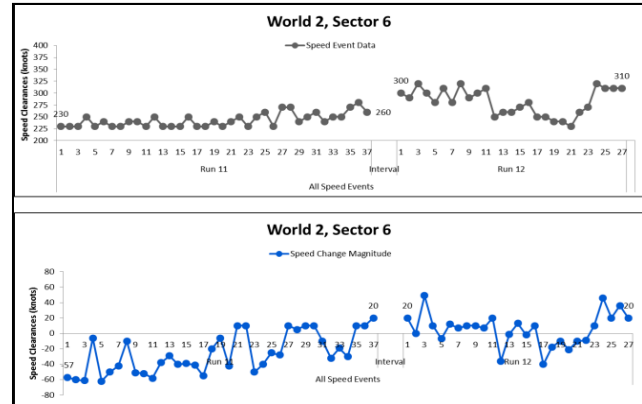

Figure 7: World 2, Sector 6's graphical display of the chronological data.

IV. DISCUSSION

Discussion of this data centers on the individual air traffic controllers' responses to the change in wind forecast error. The primary goal was to investigate whether an interrupted time-series analysis would prove beneficial in understanding

air traffic controller behavior. While this analysis does not lend to a direct comparison between air traffic controllers, examining the individual behavior does highlight potential trends. Of the data presented, speed-change magnitude is possibly more interesting, as this highlights the controller's clearances in relation to their previous decisions, allowing a visual assessment of their strategy as they adjusted their speed clearances over time. Both data types together, the raw data (speed events), and speed-change magnitude, provide a more complete picture of the controller's behavior. The analyses presented here also provide the opportunities for discussion and refinement of its application to the air traffic control domain.

The concept of level change, and whether or not their initial clearance after the interval was predictable based on their clearance strategy for run 11 (analogous to a previous shift), reveals if the controllers carried over their strategy despite the interval. World 1, Sector 5's (W1S5) level change provided insight into a controller's response to the interval when both speed events and speed change magnitude are looked at in conjunction. The fist clearance given by W1S5 after the interval fell in the predicted range (no level change), suggesting no change in strategy. However, their change in magnitude from the previous aircraft was significantly different from predicted – notably changing from +11 to -9, instead of incrementing in a positive direction like the end of run 11, suggesting the controller reversed the direction of their strategy, a pattern that continued for multiple following clearances. This behavior falls in line with the bias seen in the delay information displayed to the controllers: the underlying trajectory predictions assumed the aircraft were traveling more quickly in run 11, and less quickly in run 12. W1S5 adjusted their initial speed clearances downward as they explored the automation's error and sector conditions. W1S6, W2S5, and W2S6 all had significant level changes between runs 11 and 12, suggesting that they adjusted their strategy post-interval. This could indicate a 'reset' of their expectations about tool accuracy when approaching a new condition. Slope, which was not present in all cases, indicates a controller adjusting their clearances as they received more data. W1S6 and W2S6 both showed significant changes in their slope for speed-change magnitude during run 12, potentially indicating they were adjusting their strategy throughout the run. A visual assessment of the data shows that some controllers incremented their clearances, progressively increasing or decreasing as they learned, while some made more drastic choices.

Additional visual inspection of the data begs further analysis of the variance between clearances, something outside of the bounds of the current analysis. During the first 3 clearances of run 11, W1S5 made large jumps, a pattern also observed in W1S6 and in W1S5 (though only for speed-change magnitude). This seems to be similar to an archer gauging the distance to a target, where the first two arrows under and overshoot, with the third arrow reaching its destination. Another analogy could be that of the goldilocks principle – the bed is too long, too short, and then 'just right'. To some extent, the controllers maintained this pattern into run 12, after the interval, revealing more about the methods controllers use to gauge conditions in their sectors than about their use of a particular tool. They may be running a goldilocks test on both the sector conditions and the behavior of their system's automation. Visual inspection, combined with the presence of some significant slope changes also suggest run 12 (the negative forecast bias), may have required more refinement by the controllers to hit the 'just right' initial speed clearance. Again, further analysis of the data's variability will prove beneficial in this respect.

V. CONCLUSION

The interrupted time-series analysis (ITS), normally an experimental tool for the medical field, has potential to benefit the air traffic control research community. ITS is proving capable of providing unique insight into the cooperation between humans and automation not available through standard methods (such as averages). As noted in the discussion, level change and slope change, while useful, do not tell the full story. Using an ITS analysis revealed that an additional analysis of the progressive behavior's variance could gather actionable insight for tool design or training. Additional thought should be given to an analysis of learning, and the impact this ITS perspective could have on a human-automation teaming.

Nevertheless, the current analysis, if only by visual inspection, does lend credence to the conclusion that the controllers responded differently to the varying forecast conditions between the two runs. The speed-change magnitude data especially highlights the slope and variability of the decision-making process as they assessed the wind bias post-interval. This initial ITS analysis indicates the following: First, the controllers were responding differently between the two conditions to compensate for the forecast error, and with individual methods. Secondly, controllers are likely to have strategies for assessing the conditions of their sector and the behavior of their system's automation independent of the current state of the data. Lastly, an ITS analysis provides a unique look into air traffic control data that opens new avenues for further analysis and thought.

VI. REFERENCES

[1] Federal Aviation Administration (2013). *FAA Aerospace Forecast, Fiscal Years 2013-2033.* FAA, Washington, D.C.

[2] Federal Aviation Administration (2013). *NextGen Implementation Plan 2013.* FAA, Washington, D.C.

[3] Joint Planning and Development Office (2010). *Concept of Operations for the Next Generation Air Transportation System,* Version 3.2. JPDO, Washington, D.C.

[4] Mercer, J., N. Bienert, A. Gomez, S. Hunt, J. Kraut, L. Martin, et al. (2013). *The Impact of Trajectory Prediction Uncertainty on Air Traffic Controller Performance and Acceptability.* AIAA AIAA-2013-

4320, American Institute of Aeronautics and Astronautics, Reston, VA.

[5] Prevôt, T., N. Smith, E. Palmer, T. Callantine, P. Lee, J. Mercer, et al. (2014). *An Overview of Current Capabilities and Research Activities in the Airspace Operations Laboratory at NASA Ames Research Center*. AIAA-2014-2860, American Institute of Aeronautics and Astronautics, Reston, VA.

[6] Bienert, N., J. Mercer, J. Homola, S. Morey, and T. Prevot (2014). Case Study: Influences of Uncertainties and Traffic Scenario Difficulties in a Human-in-the-Loop Simulation. *Proceedings of the 6th International Conference on Research in Air Transportation (ICRAT)*, Istanbul, Turkey.

[7] Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (Vol. 608). John Wiley & Sons.